Claims

What is claimed is:

1. A server comprising:
a processor;
a memory;
a system area network connection;
a local area network connection; and
software operable on the processor to:
load unique content into the memory from a storage location,
receive requests for content over the local area network,
service requests for the content in memory,
service requests for content located in a memory of another server by obtaining the content over the system area network, and
cache content used to service request for content located in the memory of the other server for use in servicing subsequent requests for identical content.

2. The server of claim 1, wherein the software operable on the processor is a component of an operating system of the server.

3. The server of claim 1, wherein the software operable on the processor is a driver.

4. The server of claim 1, wherein the software operable on the processor is a middleware component.

5. The server of claim 1, wherein the system area network is a Gigabit Ethernet network.

6. The server of claim 5, wherein the Gigabit Ethernet network is TCP Offload Engine enabled.

7. The server of claim 1, wherein the unique content is loaded into memory prior to the server being available to service content requests.

8. A system comprising:

a networking logic device;

two or more servers, each server operatively coupled to the networking logic device; and

an operating system on each of the two or more servers including one or more software components including executable instructions to:

make the content of each server memory available to the other servers over the networking logic device, and

receive and fulfill content requests over the networking logic device with content from a server's local memory or from a memory of another server over the networking logic device.

9. The system of claim 8, wherein the networking logic device performs one or more of the functions of devices including:

a router;

a switch;

a firewall;

a load balancer; or

a content director.

10. The system of claim 8, the system further comprising:

an electronic storage medium system operatively coupled to the networking logic device, wherein content is stored in the electronic storage medium system.

11. The system of claim 10, wherein the electronic storage medium system is a storage area network.

12. The system of claim 8, wherein the software further includes executable instructions to cache content of other servers in memory for use in fulfilling later requests for the content in the cache.

13. The system of claim 8, wherein the software further includes executable instructions to:

maintain a table of content available on the system area network and the location of the content, and

obtain content from another server based on the table of content available on the system area network.

14. The system of claim 13, wherein the networking logic device maintains a table of content available on the system area network and routes requests based on the table.

15. A method of server operation comprising:

priming a memory of a server, wherein the server is a member of a server cluster, wherein the content in the memory of the server is unique to the server amongst all servers in the server cluster;

making the content in the server memory available to other servers in the server cluster over a high-speed interconnection;

receiving requests for content;

fulfilling content requests by retrieving data from the server memory and from memories of one or more other servers over the high-speed interconnection; and

caching content of other servers that has been requested either recently or commonly to provide the server the ability to fulfill requests for cached content locally.

16. The method of claim 15, wherein requests for content are received over a local area network connection.

17. The method of claim 16, wherein requests are received into the local area network on a router coupled to the Internet.

18. A method comprising:

distributing web content across a cluster of web servers connected by a first network;

fetching, by a first one of the web servers, web content on demand from a second one of the web servers in the cluster of web servers across the first network; and

caching the web content in the memory of the first one of the web servers.

19. The method of claim 18 wherein the web content is fetched from the memory of the second one of the web servers.

20. The method of claim 18 further comprising responding to the request with the web content from the memory of the first server.

21. An article comprising a computer-readable medium containing associated information, wherein the information, when accessed, results in a machine performing:

receiving, by a first server in a plurality of interconnected servers, a request for content; and

determining if the content is available in a memory of the first server:

if the content is available in the memory of the first server, then responding to the request with the content from the memory of the first server;

if the content is not available in the memory of the first server, then obtaining the content from a memory of one of the servers in the plurality of interconnected servers other than the first server and replicating the content in the memory of the first server.

22. The article of claim 21 further comprising responding to the request with the content from the memory of the first server.

23. The article of claim 22 further comprising responding to a subsequent request for the content with the content from the memory of the first server.